

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

LACIE-00622

JSC-11677

LARGE AREA CROP INVENTORY EXPERIMENT (LACIE)



7.9-100.49.

TM-79916

"Made available under NASA sponsorship
in the interest of early and wide dis-
semination of Earth Resources Survey
program information and without liability
for any use made thereof."

NASA NOAA USDA

(E79-10049) LARGE AREA CROP INVENTORY
EXPERIMENT (LACIE). TEST AND EVALUATION
PLAN FOR KSU AND CCEA PHASE 3 YIELD MODELS
(NASA) 16 p HC A02/MF A01 CSCL 02C

N79-13456

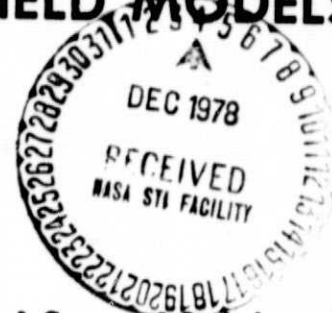
Unclass
00049

G3/43

TEST AND EVALUATION PLAN

FOR

KSU AND CCEA PHASE III YIELD MODELS



National Aeronautics and Space Administration
LYNDON B. JOHNSON SPACE CENTER

Houston, Texas

APRIL 1977

TEST AND EVALUATION PLAN
FOR
KSU AND CCEA PHASE III YIELD MODELS

Written by:

T. L. Barnett
Dr. T. L. Barnett

Approved by:

Jon D. Erickson
Jon D. Erickson
Chief, Research, Test and Evaluation Branch

Concurred by:

J. L. Rogers
J. L. Rogers
Acting Manager
Yield Estimation Subsystem

Approved by:

J. L. Dragg
J. L. Dragg
Chief
Applications Systems Verification Branch

~~PRECEDING~~ PRECEDING PAGE BLANK NOT FILMED
~~PRECEDING~~ PRECEDING PAGE BLANK NOT FILMED CONTENTS

Section	Page
ACRONYMS	iv
SUMMARY	1
INTRODUCTION	2
TASK OBJECTIVES	3
EVALUATION CRITERIA	3
APPROACH	4
DATA REQUIREMENTS	7
KSU Models	8
OCEA Models	8
Statistical Tests	8
RESOURCES REQUIRED	9
SCHEDULE	9
APPENDIX	10

ACRONYMS

CCEA: Center for Climatic and Environmental Assessment
CRD: Crop reporting district
CV: Coefficient of variation
JSC: NASA Lyndon B. Johnson Space Center
KSU: Kansas State University
LEC: Lockheed Electronics Company
NASA: National Aeronautics and Space Administration
NOAA: National Oceanic and Atmospheric Administration
SRS: Statistical Reporting Service
USDA: U.S. Department of Agriculture

TEST AND EVALUATION PLAN FOR KSU AND CCEA PHASE III YIELD MODELS

SUMMARY

The four tasks covered by the yield model test plan are as follows.

1. Evaluation of Kansas State University (KSU) yield models for Kansas, North Dakota, and three U.S.S.R. oblasts by means of 10-year bootstrap tests in Kansas and North Dakota by crop reporting district (CRD) using plot model coefficients from all available years with CRD yield adjustments calculated on the basis of the years previous to the test year; 10-year jackknife tests performed in three U.S.S.R. oblasts.
2. Comparative test and evaluation of KSU and the Center for Climatic and Environmental Assessment (CCEA) yield models for Kansas, North Dakota, and three U.S.S.R. oblasts: 10-year bootstrap tests for the KSU and CCEA Kansas, and North Dakota models, CRD by CRD as indicated above, but with trend effects prespecified; 10-year jackknife tests performed for the KSU and CCEA U.S.S.R. oblast models. Emphasis is on comparing performance of the two models.
3. Evaluation of CCEA foreign yield models: The 10-year bootstrap tests performed on models for which sufficient yield/weather historical data exist: Australia, Argentina, and Canada. Ten-year jackknife tests performed for other models: Brazil, India, China, and the U.S.S.R.¹
4. Evaluation of the CCEA Australia/Argentina model form development of the Australia/Argentina model form in Kansas: 10-year bootstrap test in Kansas.

Tests and evaluations for the above tasks, including reports and recommendations, will be complete by March 1, 1977.

¹Testing for the India and China models may be delayed beyond schedule due to lack of data and/or difficulties in developing models. Testing as described will be carried out upon delivery by the CCEA of these models.

INTRODUCTION

Several related tasks are covered under this test plan. All tasks require the same kinds of computer runs by KSU and CCEA, and all tasks must be accomplished on a similar schedule. It makes sense to run and to analyze the tests at the same time.

Task 1: Evaluation of KSU yield models for Kansas, North Dakota, and three U.S.S.R. oblasts. The KSU yield models for Kansas, North Dakota, and the three U.S.S.R. oblasts (Tselinograd, Kurgan, and Khmel-Nitsky) will be delivered in final form by December 1, 1976. The KSU model form consists of coefficients for plot-yield models based on selected weather-related variables, which apply to the entire Great Plains winter- or spring-wheat regions, and of CRD-specific adjustments of plot yield to CRD yield. Similar models for the U.S.S.R. oblasts use the appropriate Great Plains plot-yield models and oblast-specific adjustments. These models must be carefully tested and evaluated because they represent the first advanced yield models in the LACIE program.

Task 2: Comparative test and evaluation of KSU and CCEA yield models for Kansas, North Dakota, and the three U.S.S.R. oblasts. Since the KSU models are being considered as replacements for the CCEA yield models for the above areas in April 1977 (part way through Phase III), the performance of the two models must be carefully compared. There is clearly some overlap with Task 1.

Task 3: Evaluation of the CCEA foreign yield models. The CCEA yield models for the LACIE foreign areas (operational in Phase III) must be tested and evaluated. Where possible, their compliance with the LACIE 90/90 criterion (90 percent accuracy 90 percent of the time in the final LACIE production estimate at the national level) must be tested. The foreign countries involved are Canada, the U.S.S.R., Australia, Argentina, Brazil, and China. The CCEA Great Plains models should also be reevaluated if any changes have been made in them since the beginning of Phase II.

Task 4: Evaluation of the CCEA Australia/Argentina Model Form. The Phase III Australia and Argentina models represent a model form different from that used in the Great Plains. A parameter based on the Palmer Drought Index is used as an indicator of soil moisture, which replaces precipitation. Because the new model form is to go directly into operational use for Australia and Argentina, its form needs to be evaluated against the precipitation-based form in the U.S. Great Plains.

The need is evident for extensive support by KSU and CCEA in making the required model runs on computers to produce results used for test data. Support requirements are detailed later in the Test Plan.

A working group composed of members of the National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), and U.S. Department of Agriculture (USDA) has collaborated on the design of

the tests described here. A test evaluation committee chaired by a NOAA representative and consisting of representatives from the USDA and NASA will evaluate the test results and recommend model implementation or alternative action to the Project Management.

It should be noted that in the United States, tests will be conducted on a CRD basis, even though plans call for application in Phase III on a partition basis. (A "partition" is an area that is relatively uniform according to various criteria, mainly soil type; partition boundaries, however, are not constrained to political boundaries). The reasons for using a CRD basis are that (1) the partitions have been defined only recently for the first time and will undoubtedly change at least once before April 1977, and (2) yields are available historically for CRD's, and techniques for allocating historical yields among partitions have yet to be defined.

TASK OBJECTIVES

Task 1: KSU Model Evaluation. To establish that the LACIE 90/90 criterion is met; to evaluate the performance of the KSU model, particularly the accuracy and responsiveness to other-than-normal weather conditions.

Task 2: KSU/CCEA Model Comparison. To compare performances of the two models in the regions of their applicability.

Task 3: CCEA Foreign Model Evaluation. To establish that the LACIE 90/90 criterion is met, or to establish the actual accuracy of each CCEA foreign model; to evaluate performance of each model, particularly the accuracy of prediction and response to other-than-normal weather conditions.

Task 4: Australia/Argentina Model Form Evaluation. To compare performances of the Australia/Argentina model form developed specifically for Kansas with that of the CCEA operational Kansas model.

EVALUATION CRITERIA

Evaluating the performance of a yield model will be based primarily on the following criteria:

1. Compliance with the LACIE 90/90 criterion or nearness to it.
2. Response to other-than-normal weather.

3. Mean level bias, $\bar{b}$, over the test years.

$$\bar{b} = \frac{1}{n} \sum_{i=1}^n (Y_{\text{pred},i} - Y_{\text{SRS},i})$$

4. Root mean square error, $(\sqrt{\text{MSE}})_y$, over the test years.

$$(\sqrt{\text{MSE}})_y = \left\{ \frac{1}{n} \sum_{i=1}^n (Y_{\text{pred},i} - Y_{\text{SRS},i})^2 \right\}^{1/2}$$

The specific tests and displays of test results that were designed to permit evaluation using the above criteria are discussed in the next section.

APPROACH

The basic problem in testing yield models is to find sufficient test data, or alternatively, to find the best test within the constraints of available data. The number of available years of data (n) at a location is rarely enough to permit 10 years to be held apart and still allow the remaining ($n - 10$) years to represent adequately the effects of weather on yield. This problem exists particularly in the CCEA models.

The CCEA yield models are built by direct regression of state historical reported yields using variables based on state-averaged/monthly-averaged maximum and minimum temperatures and state-averaged monthly precipitation totals. Trend terms are specified by fitting a series of straight lines to the historical yields over the years. A characteristic of the CCEA models, the importance of which will become evident later, is that these models specify the trend effects based on all historical years up through 1975. The 1976 prediction involves an extrapolation of the trend, but tests on historical years become interpolations within a prespecified trend line. In the U.S. Great Plains, the models are built on state data but are in practice applied to each CRD using CRD-averaged weather data. In foreign countries, the models are built using data from crop-reporting regions roughly corresponding to states.

The KSU models are developed from plot yields and daily weather observations at branch experiment stations throughout the Great Plains. Over 1000 observations went into building the winter-wheat plot-yield model, and a similar number of observations will go into the spring-wheat model. As a result, the winter- and spring-wheat KSU plot-yield models are "universal" for the Great Plains, and should also apply to foreign areas without modifications.

To produce a CRD-based yield model, the plot-yield model predictions are adjusted using a ratio to best match the Statistical Reporting Service (SRS) reported yields on a CRD basis. The process is performed in a similar manner for oblasts, provinces, and so forth. Therefore, the KSU model consists of a set of "universal" coefficients of weather-related variables and one yield-adjustment term for each CRD, oblast, or other region to which the model applies.

The testing process must approximate as nearly as possible the way the models will be used operationally in predicting next year's yields. Not only must the tests of the KSU and CCEA models be the best possible considering the constraints of the available data, but they must give results which are directly comparable. This comparability is hard to achieve because of the different natures of the two models. Coefficient stability in the KSU model appears to be high enough that a true 10-year independent test - that is, holding out the last 10 years and developing coefficients only on the years up to $(n - 10)$ - could be carried out. This is not possible, however, with the CCEA model. Therefore, the best test that can be run for the CCEA model necessarily sets the pattern for both model tests.

Another factor that makes comparability particularly difficult is the different ways in which the trend effects are handled. The KSU model attempts to model the trend effects from fertilizer usage, from variety mix, and from fallow practices, whereas the CCEA model specifies trend empirically from the analysis of all available years. In our tests under consideration, the latter practice has had the effect of prespecifying trend effects and, therefore, of removing them from consideration. Root mean square errors for predictions are those caused by weather only. Unless one postulates that he "knows" the trend next year is going to follow the trend over the past several years, the root mean square error contribution resulting from the uncertainty involved in projecting the trend from one year to next year is missed. To insist that CCEA recalculate the trend annually, as they did in Phase-I models, is not fair either because this practice introduces an additional important degree of freedom not found in the operational model. The root mean square error resulting from such a procedure would not represent a reliable estimate of the error in predicting the yield for the following year.

It appears that the fairest way of comparing the KSU and CCEA models would be to effectively prespecify the trend effects as much as possible in the KSU model as well. This can be done by using the fertilizer-usage model over all the available years (or the real fertilizer amounts) and inserting the real-variety mix and the fallow practices for each test year. Notice, however, that this set of test results should not be used in Task 1 (the evaluation of the KSU model). The proper test would involve estimation of fertilizer usage, variety mix, and fallow practices for a test year only from past years.

The best approach in testing the models to satisfy the four tasks appears to be the one described below. The tests simulate as closely as possible the situation involved in predicting each year's yield from models developed on all available data to date.

In situations where about 20 or more years of historical yield/weather data are available, the "bootstrap" approach is used. In a 10-year bootstrap test, the model coefficients are developed only from the years up through $(n - 11)$ and are used to predict the yield for year $(n - 10)$. Then the model coefficients are rederived from the years through $(n - 10)$ and are used to predict the yield for year $(n - 9)$ and so forth. In this manner, 10 years of relatively independent tests results are produced to serve as test data.

In situations where less than about 20 years of yield/weather data are available for a region, a somewhat different approach, the "jackknife" approach, must be used. The $(n - 10)$ years are insufficient to develop a reliable set of model coefficients. Therefore, one year is held apart; the model coefficients are developed for the remaining $(n - 1)$ years, and tests are conducted on the held-out year. The held-out year is then permuted among the n available years, or among a specified subset. In particular, for a 10-year jackknife test, one permutes the held-out year among the last 10 years, but in every case develops the coefficients from the $(n - 1)$ years. The main drawback to the jackknife test, as opposed to the bootstrap test, is that the trend terms, if not already prespecified as in the CCEA model, are found by interpolation rather than extrapolation. This process may make the test somewhat less rigorous in predicting the performance of a model.

The available years of data for each model to be tested, and the testing approach dictated, are listed below.

Model	Years	Test
U.S. Great Plains	1920's-1975	10 year bootstrap
Canada	1930's-1974	10 year bootstrap
U.S.S.R.	1958-1972	10 year jackknife
Australia	1940-1972	10 year bootstrap
Argentina	1931-1970	10 year bootstrap
Brazil	1962-1973	10 year jackknife
India	1956-1965	10 year jackknife ^a
China	?	10 year jackknife ^a

^aThe exact data available and model readiness dates for the India and China models are not yet clear. The models will be tested when delivered. Tests will correspond to those described as nearly as possible.

Models will be tested at the same regional level at which they are used in LACIE. This means CRD (with aggregation to state level) in the United States, province in Canada, oblast in the U.S.S.R., and appropriate crop reporting regions in the other countries.

The KSU will run test results for the KSU models. They are best equipped to do so on a quick turnaround and will be essentially free to do so beginning in December. The KSU will use daily historical weather data from the Kansas and North Dakota branch experiment stations as their test data base. Although the weather data is not the same as in the CCEA first-order met station data base, the station density (one per CRD) is similar. The KSU data base is considered to be entirely adequate because a point model is being compared on the basis of daily weather input against a state model based on monthly state averages. The 10-year bootstrap tests in each of the Kansas and North Dakota CRD's will be run with a plot-yield model based on all available plot-weather/yield data. The plot-weather/yield data constitutes a data set as nearly independent of the met station weather and CRD yields as is probable in practice. The daily historical oblast yield data for the three U.S.S.R. oblasts will be supplied to KSU by CCEA. The CCEA will run the required tests for the CCEA models. These models consist of CRD-by-CRD runs for Kansas and North Dakota using 10-year bootstrap tests, region-by-region runs for each LACIE foreign county using 10-year bootstrap or 10-year jackknife tests, development in Kansas of a model of the Australian/Argentina form, and running a 10-year bootstrap test, CRD-by-CRD and statewide, to compare this model to the operational CCEA Kansas model.

Using the test results supplied by KSU and CCEA, Lockheed Electronics Company (LEC) statisticians at the NASA Lyndon B. Johnson Space Center (JSC) will compile the following statistical parameters and displays for each CRD and state in the United States and for each appropriate crop reporting region in foreign countries. They will consider the following data: The bias b over the test years; the root mean square error, $(\sqrt{MSE})_y$; the plot of predicted yields for the two models and SRS yields versus years; and the plot of predicted yield versus the SRS yield over the years for each model.

The last two considerations indicate qualitatively how well a model responds to other-than-normal weather, and the deviation of the least squares linear fit in item 4 from the ideal 1:1 line indicates the degree to which the model over- or under-responds. Response to other-than-normal weather, or unusual relationships not monitored well by a model, can be studied by looking at the worst year for each model, the second worst year, and so forth in an attempt to identify the type of non-normal crop developments, weather patterns, diseases, and other abnormalities that the model fails to monitor.

DATA REQUIREMENTS

Data required to run the tests are as follows.

KSU Models

The KSU models (to be run at KSU) will include the following: SRS final yield figures by CRD for Kansas and North Dakota for 1948 (or earlier) through 1975 (on hand at KSU); daily historical weather ($T_{\max}$, $T_{\min}$, precipitation) at branch experiment stations in Kansas and North Dakota for same years (on hand at KSU); official U.S.S.R. yields for Tselinograd, Kurgan, and Khmel-Nitsky oblasts for 1958 through 1972³ (to be supplied to KSU by CCEA); and daily historical weather ($T_{\max}$, $T_{\min}$, precipitation) from first order met stations in and near the test oblasts for same years (to be supplied to KSU by CCEA).

CCEA Models

The CCEA models (to be run at CCEA) will include yield and monthly historical weather data from which CCEA models were developed for each state and each foreign crop reporting region³ and monthly historical weather data for each CRD in Kansas and North Dakota (on hand at CCEA).

Statistical Tests

Data required to run the statistical tests (results of the bootstrap or jackknife runs of yield models) are as follows.

1. Predicted yearly yield, model parameter values, and independent data values for 10-year bootstrap runs of KSU models for each CRD in Kansas and North Dakota. These data must be run with trend effects prespecified (that is, using the fertilizer-usage model from all years and the real-variety mix and fallow practices for each test year) for use in Task 2 (to be supplied by KSU).
2. Same as item 1 except that the run is made with the fertilizer model developed only from years previous to each test year and that use is made of projections about variety mix and fallow practices for each test year for use in Task 1 (to be supplied by KSU).
3. Predicted yearly yield, model parameter values, and independent data values for 10-year jackknife runs of KSU models for three U.S.S.R. oblasts (to be supplied by KSU).
4. Predicted yearly yields, model parameter values, and independent data values for 10-year bootstrap runs of CCEA models for each CRD in Kansas and North Dakota and for states (to be supplied by CCEA).

³And later years if available.

5. Predicted yearly yields, model parameter values, and independent data values from 10-year bootstrap or jackknife runs, as appropriate, for crop reporting regions of each LACIE foreign country (to be supplied by CCEA).

6. Model developed on Kansas using the Australian/Argentinian model forms and predicted yearly yields, model parameter values, and independent data values from 10-year bootstrap runs in each Kansas CRD and for the entire state (to be supplied by CCEA).

7. Predicted yearly yields, model parameters values, and independent data values from 10-year bootstrap runs in each CRD and state for any United States models that have changed significantly from those evaluated for Phase II operations (to be supplied by CCEA).

RESOURCES REQUIRED

Estimates of manpower (Civil Service and contractor) required to accomplish the testing and evaluation jobs outlined here are as described below. Additional funding is not required.

Running models required to provide test data at KSU are estimated to require 2 man months (MM); at CCEA, 12 MM.

The cost of reducing and analyzing test results at JSC is estimated to require 8 MM using LEC personnel.

The costs involved in evaluating test results is divided between the costs for working group members required for the task, which is estimated at 3 MM, and LEC support, which is estimated at 2 MM.

Documentation of results, which would be a function of LEC support, is estimated at 2 MM.

SCHEDULE

The tasks defined in this plan will be performed in accordance with the schedule presented in Table 1.

APPENDIX

TEST OF COMPLIANCE WITH LACIE 90/90 CRITERION

Compliance tests for the CCEA models have, to date, been conducted as follows. To determine whether a state yield model meets the LACIE 90/90 criterion at the national level in production, one considers the probability statement

$$\Pr \left(\frac{|\hat{P}_1 - P_s|}{P_s} \leq \frac{0.0707}{\sqrt{\delta_s}} \right) \geq 0.90$$

where P_s is the state production (spring or winter wheat as appropriate); $\hat{P}_1 = A \times \hat{Y}$ and A is the SRS acreage for the state; $\hat{Y}$ is the model estimated yield; δ_s is the fraction of national production contributed by the state, oblast, or other location in question; and the factor 0.0707 arises from the product of 0.10, which corresponds to a 10-percent error (90 percent accuracy) times $\sqrt{2}$ arising from the assumption that the error is equally distributed between acreage and yield. The assumptions implicit in the above are (1) $\hat{P}_1$ is an unbiased estimator of P_s and is normally distributed; (2) $\delta_s = P_s / P$, where $P = \text{var}(\hat{P}_s) / \text{var}(\hat{P})$, where P is the total national production; and (3) $\text{var}(\hat{P}_s) = 2 \text{var}(\hat{P}_1)$.

To devise a test for determining whether the probability statement above is satisfied, the following procedure is proposed.

$$\psi(x) = \begin{cases} 1 & \text{for } x \leq 0 \\ 0 & \text{for } x > 0 \end{cases}$$

and define the random variable.

$$z = \frac{|\hat{P}_1 - P|}{\sqrt{\delta_s}} - \frac{0.0707 P}{\sqrt{\delta_s}}$$

Then test the null hypothesis.

$$H_0: E(\psi(Z)) \geq 0.90.$$

If $z(t)$, $t = 1, 2, \dots$ denote independent observations of Z , then the test can be based on the following binomially distributed statistic.

$$\hat{p}_n = \frac{1}{n} \sum_{t=1}^n \psi(z(t))$$

Then let

$$\pi_s = \Pr \left(|\hat{p}_1 - P| / P \leq \frac{0.0707}{\sqrt{\delta_s}} \right)$$

Tests of the 90/90 criterion at the state level reduce the form of the tests to

$$H_0: \pi_s \geq 0.90 \text{ versus } H_a: \pi_s < 0.90.$$

The statistic for this test is $\hat{p}_n$. Using the fact that $n \hat{p}_n$ has a binomial distribution, a one-tail test of H_0 at the 0.07 level for $n = 10$ is to accept H_0 if

$$10 \hat{p}_{10} \geq 8;$$

Otherwise, reject.

Now that estimates of LACIE acreage estimate errors are available, they should be used in evaluating compliance with the 90/90 criterion. To do so, the coefficient of variation (CV) for production is computed for each state using the yield errors resulting from the 10 years of tests and the acreage errors estimated in the latest available year (presumably 1975-76). In the case of the U.S.S.R. oblasts, it will be necessary to compare 1975 and 1976 acreage estimates to the latest available reports, probably 1974. The CV for a state is computed by dividing the square root as the variance in production for the state by the actual (true) production in the state. The variance in

production should be calculated from both state yield estimates and, if available, from CRD yield and acreage estimates aggregated according to LACIE guidelines up to the state level. The resulting CV's must be scaled to the national level and compared to the project goal of 6 percent, which corresponds to the 90/90 criterion.